

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----X
UNITED STATES OF AMERICA :

- v. - : 15 Cr. 565 (VEC)

KEVIN JOHNSON, :

DEFENDANT. :

-----X

**REPLY MEMORANDUM OF LAW IN SUPPORT OF
KEVIN JOHNSON'S MOTION TO EXCLUDE EVIDENCE AT TRIAL
GENERATED BY, AND TESTIMONY ABOUT,
THE FORENSIC STATISTICAL TOOL (FST)**

DATED: New York, New York
February 16, 2017

Respectfully submitted,
DAVID E. PATTON, ESQ.
Federal Defenders of New York

By: /s/
CHRISTOPHER FLOOD, ESQ.
SYLVIE LEVINE, ESQ.
ROBERT M. BAUM, ESQ.
Attorneys for Defendant
Kevin Johnson
52 Duane Street - 10th Floor
New York, New York 10007

TO: PREET BHARARA, ESQ.
United States Attorney
Southern District of New York
One St. Andrew's Plaza
New York, New York 10007

Attn: **STEPHANIE LAKE, ESQ.**
SHEB SWETT, ESQ.
Assistant United States Attorney

Table of Contents

I.	Introduction	1
II.	FST represents OCME’s defection from the legal and scientific compact in place since first NRC Report.	1
A.	The hard-earned balance between law and science.....	1
1.	Working to establish the primacy of wisdom over novelty.	1
2.	OCME adopts rules known only to itself.	4
3.	The OCME adopts definitions of material terms known only to itself.	6
B.	Daubert: Clear guidance from a flexible standard.	9
C.	FST poses an unacceptably high risk of false positives and exaggerated false associations.	13
D.	FST causes new constitutional harm in the guise of avoiding another.	14
III.	The Government fails to demonstrate that FST is admissible.	16
A.	The record of FST’s scientifically unreliable defects and limitations.	16
1.	Issues the Government concedes.....	17
a.	The formula of FST obtained under Protective Order is not the formula subject to the validation study.....	17
b.	Locus-dropping behavior remains undocumented, unexplained.....	18
2.	Implications of FST being employed without validation or approval.	19
a.	Eliminating data is only helpful to a hypothetical contributor.	19
b.	Conflation with NRC II <i>theta</i> is misleading.	20
c.	Discredited hand calculations have no relationship to dropped alleles.....	20
3.	Documenting how hidden behavior eliminates data and the effect on the LR....	21
a.	The drop-out rate was not empirically determined.	23
4.	The Government misunderstands the science.....	23
a.	The degradation module did not work.....	23
b.	Threshold obscures valuable data.....	24
5.	The Subcommittee was an inadequate stop gap.....	25
a.	Simulated alleles were obscured from the Subcommittee.	25
IV.	The Government’s <i>Daubert</i> factor analysis fails to support FST’s admissibility	26
A.	Factor One: Testing.....	26

B. Factor Two: Peer review..... 26

C. Factor Three: Known or potential error rate. 28

D. Factor Four: Protocols and review. 29

E. Factor Five: Scientific community.31

V. Conclusion 32

I. Introduction

The New York City Office of the Chief Medical Examiner's (OCME) software Forensic Statistical Tool (FST), used in this case to produce two Likelihood Ratios (LR) purporting to associate Mr. Kevin Johnson with contraband evidence is not scientifically valid. For reasons stated previously, as well as those stated below, and including any that the Court may so find, we respectfully submit that it cannot help the trier of fact to discern or understand relevant facts or evidence, or to determine a fact in issue. Mr. Kevin Johnson respectfully seeks a ruling *in limine* excluding any and all evidence generated by, and testimony relating to, FST. Pursuant to an Order issued on January 6, 2017¹, the Court has already ordered the parties to appear for an evidentiary hearing on this motion.

On facts where a public forensic laboratory has repeatedly shown contempt for the public's right to information, where that forensic laboratory has unilaterally and without notice elected to alter the terms by which forensic science and the law have coexisted for decades, the Government does not contest and chooses not to explain the OCME's most perplexing misconduct. Rather, it invites the Court to join in the lab's adventure by adopting a novel, entirely permissive standard that reimagines Daubert and the Federal Rules of Evidence (hereinafter "Rules") as welcoming all but "junk science." We may well prevail even under the Government's creative standard, but that is not the law. Its invitation should be rejected.

II. FST represents OCME's defection from the legal and scientific compact in place since first NRC Report.

A. The hard-earned balance between law and science.

1. Working to establish the primacy of wisdom over novelty.

Courts decide, not scientists, how tension is to be resolved between novel concepts and the fair, orderly, and just administration of law.² Best-situated to strike the appropriate balance of rights and interests and empowered to decide, in the three decades since forensic DNA typing has been a fixture in criminal investigations, courts have worked to define the conditions by which it may be admitted into evidence. Abundant resources extend the body of case law: National guidelines chartered by the White House³, fostered by the Federal Bureau of Investigation

¹ See Docket No. 96, Order.

² *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 597-98 (1993).

³ Executive Office of the President's Council of Advisors on Science and Technology ("PCAST"), Def. Br., Ex. L.

(FBI)⁴, and the National Institute of Standards and Technology⁵; extensive scholarship⁶; and persuasive secondary source material.⁷ Perhaps most significant among them are the seminal commissioned reports of the National Research Council (NRC)⁸.

The NRC I committee undertook a comprehensive study of DNA typing. The eminent Johns Hopkins Hospital medical geneticist and Professor of Medicine, Dr. Victor McKusick chaired, organizing a multidisciplinary committee of experts and scholars, philosophers and lawyers, researchers and clinicians from different elite institutions and covering the public and private sector. The bioethicist Ruth Macklin and the Honorable Judge Jack B. Weinstein of the Eastern District of New York completed the roster of an indisputably blue ribbon committee.

The enduring structural, conceptual, and ethical contribution of NRC I was to take the emergent field with profound seriousness. In doing so, the committee established a credible basis for leadership across the disciplines of law and forensics. Bedrock precepts in DNA forensics trace their origin to NRC I: That DNA typing's components – laboratory analysis, comparison techniques, and statistical methods⁹ – are practically and conceptually inseparable.

The nation's first clear statement on the necessity of validation of DNA typing techniques – both internally and for casework – also came from the committee: Before any change to an existing, typing procedure, labs must “rigorously characterize” the method “in both research and forensic settings” to seek where it

⁴ See Scientific Working Group on DNA Analysis Methods (SWGDAM) Guidelines for the Interpretation of Autosomal DNA, Gov't Br. at 8.

⁵ Organization of Scientific Area Committees (OSAC) (part of a NIST- and DOJ-sponsored initiative), *available at* <https://www.nist.gov/topics/forensic-science/osac-home> (last visited February 15, 2017).

⁶ Andrea Roth, *Safety in Numbers? Deciding When DNA Alone is Enough to Convict*, 85 NEW YORK UNIVERSITY LAW REVIEW 1142 (2010); Erin E. Murphy, *The Mismatch Between Twenty-First-Century Forensic Evidence and Our Antiquated Criminal Justice System*, 87 S. CAL. L. REV. 633 (2014).

⁷ Erin E. Murphy, *Inside the Cell: The Dark Side of Forensic DNA* (Nation Books 2015).

⁸ See National Research Council, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992) (“NRC I”); *see also* National Research Council, THE EVALUATION OF FORENSIC DNA EVIDENCE (1996) (“NRC II”); *and see* National Research Council, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES, A PATH FORWARD (2009) (“NRC III”).

⁹ NRC I at pp. 6-8.

may be reliably employed.¹⁰

The committee was correct, even if their recommendations were not always followed. Many problems NRC I confronted remain to this day for want of implementation of the sensible proposal recommended by the committee. For example, NRC I called for mandatory federal accreditation of all forensic DNA labs. Observing that neither internal Quality-Assurance checks nor voluntary accreditation through private professional organizations like the American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLD-LAB) provide sufficient means of achieving consistent adherence to the appropriately high scientific standards.¹¹

Spanning three decades, NRC I's viewpoint has become canonical. Its methods and recommendations implicitly model principles of conservatism, good faith, and adherence to the highest standards of excellence that subsequent NRC committees dutifully made explicit. Working to the strengths of both the legal and the forensic communities it was tasked to serve, recognizing the timelessness of the one and the nimbleness of the other NRC I truly has stewarded every major development in the law as it applies to DNA typing since the earliest days of its introduction.

The advent of typing methods with previously unimaginable sensitivity has engendered the development capabilities for capturing DNA profiles from surfaces and places that previously would not have been thought to have evidential significance. But focus on such small amounts of material presents serious new challenges. Complex mixtures, partial profiles, degradation, small template amounts that behave unpredictably in the lab – separately, any one of these features has the potential to make accurate interpretation difficult. All these features are combined in Kevin Johnson's case.

Paradoxically, the more sensitive DNA becomes, the more it drains of its famed discriminatory power. How to assign weight, and what weight to assign an association between the evidence and the accused is moving to the center of the discussion. Not since the early 1990's, when the population geneticists from elite academic institutions dueled in the courts over the meaning of population substructure¹² has the need for the guidance of NRC operational ethics been greater.

¹⁰ *Id.* at p. 8.

¹¹ *See* NRC I at pp. 16-17.

¹² In that vigorous debate, intellectual heavyweights were on opposite sides: population geneticists Richard Lewontin of Harvard and Daniel Hartl of Washington University against Ranajit Chakraborty, then of the University of Texas, and Kenneth Kidd of Yale. Observers described the dispute—taking place in the pages of *Science* and elsewhere—as

But FST makes bold claims from the most questionable samples. It associates people with DNA samples of the poorest quality, the most degraded, complex mixtures – even those without complete profiles that previously would have been considered forensically useless. This raises the specter of DNA false-associations.¹³ That the very technology once heralded as a truth machine, solving cold cases and exonerating the innocent could now become a threat to falsely include bespeaks a breakdown somewhere in the system.

This only emphasizes the crucial and enduring role courts play in scrutinizing evidence. Principles articulated in the early 1990's by the first learned NRC committee relating to who must bear the risk of limited uncertainty and who benefits from statistical and mathematical conservatism still define the system's ethics. Technological details may change over time, and adversarial interests will differ, but these operational ethics first defined by NRC I presumptively still endure.

2. OCME adopts rules known only to itself.

The OCME defected from NRC I's cannon without notice. They have never made it clear why or when they did. It's now clear that that they have. It does not take very long struggling with the public lab's wholly-owned actuarial software, FST to conclude that its use is not compatible with the presumption of innocence. The values implemented in FST demonstrate that the lab is operating by a different ethic than the protection of the innocent or the very highest standards of scientific practice. Looking no further than FST's internal validation has been made proxy for proper fieldwork and it is clear that, as compared with the validation In its development, the OCME disregarded law, scientific method, and the standards governing all of forensics since the release of NRC I.

The manner in which the program was created was neither scientific nor scientifically reliable. FST's conceptual model was never shared beyond the small circle of its creators, nor were the designs, data, or testing ever exposed to outside review. Before FST could be used in casework, however, the OCME was required by law to obtain the approval of the New York State Forensic Science Commission Subcommittee on DNA ("the Subcommittee").¹⁴ In this, however, the lab thwarted what minimal oversight the Subcommittee could exercise by misrepresenting how FST would be used in casework, buttressing its application with false and exaggerated data, and misleading the Subcommittee about material aspects of the

"decidedly nasty." See Leslie Roberts, Fight Erupts Over DNA Fingerprinting, 254 Science 5039 (1991), pp. 1721-23.

¹³ *Id.*

¹⁴ N.Y. Exec. Law § 955-b.

FST's internal validation.

FST was never truly validated for use in forensic casework. At best, the accumulated binders of material cited in the Opposition Brief represents the record of a troubled first step in a validation process. It is comprised of sample sheets, electropherograms ("epg's"), tabulated observations of PCR-STR tests performed on mock evidence – a voluminous record of swabbing, mixing, and running samples, reading and recording epg results. But the design and method of the "Drop Study": recording the "rates" at which alleles drop out from samples, or randomly appear in them, did not model the qualities of actual forensic samples. That the OCME went ahead to conduct casework with FST notwithstanding the limitations of the foundational data is dangerously unreasonable, invalid, and scientifically unsound.

Beyond biological testing, the conceptual modeling of FST's LR formula evidences no consideration of the context in which it would be used. Among the intractable problems FST presents to integrating with the defense as it is currently practiced in American courts: It shifts constitutional burdens; quantifies the presumption of innocence; miscomprehends how a theory of defense might operate; forces a theory defense upon the defendant in advance; requires the adoption of features of a theory – like allelic drop-out – that the defense may reject entirely; creates a false equivalence between the prosecution and defense, and relies upon the presumption of guilt in order to function. With these barriers to integration with federal jurisprudence, it comes as little surprise that no attempt has been previously made to introduce FST over defense objection in this court. It is also clear that, at minimum, FST must be subjected to "careful scrutiny" and inquiry into its limitations, and the circumstances in which it will yield unreliable results.¹⁵

FST doesn't provide much to guide the analysis. Its reports are studies in obfuscation, providing far less detail than necessary to comprehend the nature of the calculation. No expert, layperson, jurist or juror can discern its meaning with this information alone. The attendant "qualitative scale", placing the LR within one of five categories of support for the Prosecutor's Hypothesis only exacerbates FST's evidentiary failings, appears to contradict the program's stated purpose, and has

¹⁵ As PCAST defines "probabilistic genotyping systems", however, it is evident that FST's inclusion within this category is overly generous to FST, as OCME's program lacks nearly every defining capability, and is beset by the major vice of interpreter bias. Regardless of its passing inclusion in the first PCAST report, FST has certainly earned the scrutiny PCAST recommends.

itself never been validated.¹⁶ The manner in which the OCME reports FST results bespeaks an indifference¹⁷ that borders on contemptuousness for the litigants who receive them. FST has never been bound by any ground truth, observable scale or other objective factor. Perhaps the most tangible feature of FST is the likelihood that, as it is employed in casework, it will lead to a false inclusion or falsely exaggerated associations or other material error. That FST fails to meet the baseline for forensic application, a prominent feature of the program is its risk of false inclusion.

NRC I's precept that "[a]ll data and laboratory records generated by analysis of DNA samples should be made freely available to all parties. Such access is essential for evaluating the analysis"¹⁸ has few proponents among FST's proprietors. Struggles to obtain necessary information to evaluate an adverse DNA analysis are the overwhelming norm, while in many other labs is it the notable exception. But why? The senselessness is striking. A senior official described the policy of not providing electronic data files – necessary to observe the steps taken by the analyst in testing – as a step too costly to implement. In response to the a fast offer by the defense to provide the media *every time*, the stonewalling continued as he explained that the cost issue was the administrator who had to click on the mouse.

This conduct is not isolated, but common; not accidental, but part of a culture that does not adhere to NRC I or rules inferentially related to it.

3. The OCME adopts definitions of material terms known only to itself.

It can be confidently assumed that Discovery requests would never produce, and it is highly questionable whether validation could have addressed the most radical of departures observed at the OCME. These distinguish FST as singularly unreliable, and raise troubling questions about the lab that have gone unanswered in the Opposition Brief. Those are deeply concealed, and would be impossible to detect short of extensive litigation. The version of FST currently in use has been materially

¹⁶ *Cf. United States v. Jiau*, 11 Cr. 161, 2011 WL 1810166, at *37 (S.D.N.Y. Mar. 8, 2011) (Judge Rakoff criticizes U.S.S.G. sec. 2B1.1: "The Guidelines give the mirage of something that can be obtained with arithmetic certainty.")

¹⁷ For the Office of the Inspector General's account of an incident in which a LR value exonerating the defendant was reported to the defense, but in which the defendant nevertheless entered a guilty plea, and the lab took no action despite updated favorable results, *see* Caragine, *infra*.

¹⁸ NRC I at pp. 16-17.

altered from the one that was approved by the Subcommittee in October 2010.¹⁹ The “undocumented behavior” discovered in the FST source code²⁰ affects the program’s LR by arbitrarily dropping loci from the calculation in ways that do not obey a consistent premise.²¹ This has never been documented; never been validated; and it is not denied.

In the Opposition Brief, the indefensible is labeled ‘unhelpful.’ To whom? It depends entirely on whose Ox is being gored. It’s a question of subjective perspective on the evidence, which somehow the Opposition Brief expects the Court to miss. But taking the claim seriously resolves a persistent mystery surrounding the lab’s terminology.²² Doing so reveals the depth of the OCME’s departure from NRC’s ethical framework.

The fact is, it is not possible to predict in advance what effect an equation that does not explicitly favor either the Hp or the Hd will have in a hypothetical future LR. The assertion that the “existence of such common alleles yields no *helpful* information”²³ appears at first reckless and misleading. There are too many variations in the way evidence is presented, too much uncertainty. How could the OCME believe this?

FST’s documentation abounds with equally puzzling use of the term ‘conservative.’ The terms aren’t simply being used in the same manner. They are saying virtually the same thing: It is an expression of how the OCME sees the value of the evidence.²⁴ At root, both trace directly to the OCME’s defection from the NRC

¹⁹ See Declaration of Nathaniel Adams, February 12, 2017 (“Adams Decl. II”), Attached at Exhibit A.

²⁰ See Adams Decl., Def. Br., Ex. C. at ¶ 5.6.

²¹ We demonstrate through a single test of validation study data the substantial effect this alteration of the code has on an LR value. Adams Decl. II at 3.

²² For ease of reference, unless otherwise specified, the term “documentation” is intended to encompass the two articles published: Def. Br., Ex. M (Adele A. Mitchell, Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop in, 6 Forensic Sci. Int’l: Genetics 749-761, 756 (2012) (“Mitchell et al.”)) and Adele Mitchell, et.al, *Likelihood ratio statistics for DNA mixtures allowing for drop-out and drop-in*. Forensic Sci. Int’l: Genetics 3 (August 31, 2011), as well as the testimonial record in *Collins* and *Rodriguez*, and recorded transcripts of presentations to the Subcommittee.

²³ Gov’t Br. at 50.

²⁴ It is a difficult adjustment to make, for one entirely acclimated to the NRC I paradigm, related as it is to the rest of American jurisprudence and focusing on the rights of actual people. But it is not necessarily a person who benefits in this new orthodoxy.

orthodoxy. NRC II explicitly defines the term as it is universally understood in the forensics:

Conservative – favoring the defendant. A conservative estimate is deliberately chosen to be more favorable to the defendant than the best (unbiased) estimate would be.²⁵

This is no longer the operative definition being used by the OCME. It was most succinctly stated by Teresa Caragine in response to a passing question during the presentation to the Subcommittee on October 8, 2010. She was asked to explain certain aspects of the statistical calculations. Responding, she described the meaning of the term “conservative” is to “favor the non-contributor.”²⁶

Theresa Caragine’s admission, the term constantly repeated but the definition never volunteered²⁷ is consistent with that of Adele Mitchell under cross examination during the admissibility hearing in *People v. Collins*²⁸. It is also consistent with the position taken in the Opposition Brief, that the undocumented behavior, eliminating what could otherwise be probative information – information demonstrably impactful on a LR value – is *a priori* not helpful. This position can only be adopted if one is not favoring a person, but a concept. That is a break from the ethics of the NRC, the structure that has held for over twenty five years.

Who is this hypothetical non-contributor who is built immutably into the formula? Because it is this imaginary person, not the *actual* defendant in a DNA-based prosecution, who receives the benefit of this “conservatism.” The scale of the lab’s misconduct here is difficult to understate. At issue is far more than vocabulary. The wisdom of the NRC committees that forged the bedrock principles to govern and guide forensic practices was not lightly offered or easily obtained. Conservative estimates cohere with the presumption of innocence, and the burden and standard of proof. They correctly interact with settled constitutional precedent, assigning the cost of limited uncertainty to the proponent.

Here, behind the screen of the LR’s opacity, the OCME upended NRC’s settled

²⁵ NRC II at p. 215.

²⁶ See Transcript of Theresa Caragine presentation to Subcommittee, October 8, 2010.

²⁷ Reviewing the documentation, in the dozens times OCME officials champion FST as “conservative”, not once is this fundamental departure from common usage volunteered.

²⁸ *People v. Collins & Peaks*, Ind. Nos. 80077-2010, 7690-2010 (Sup. Ct. Kings Cnty. Nov. 7, 2014) (hereinafter “*Collins*”).

framework. For this documented misconduct²⁹, the Government must bear the cost.

B. Daubert: Clear guidance from a flexible standard.

The *Daubert* opinion is remarkably lucid. Its sturdy language bespeaks care and competence in the momentous decision to overturn settled – albeit controversial³⁰ – precedent. Superseding *Frye* meant far more than affirming the traditional gatekeeping role of the district court. It signaled confidence in district courts’ ability to comprehend and manage complex questions. Fully informed³¹, the Court supplanted the rigid and “austere”³² general acceptance standard in favor of a searching inquiry that expands judicial responsibility and authority as the arbiter of

²⁹ It is noted that the Government attempts to introduce a bad faith standard from the civil arena, which is not applicable in the criminal context. We do not subject ourselves to it by observing that a perfectly serviceable operative definition of bad faith would be for one to employ a term of common usage while intending a meaning that sharply diverges from common usage but which cannot be verified, one that the listener neither shares nor would see in his or her personal interest, and failing to resolve the ambiguity either by providing notice, or by returning to conventional use.

One party in the exchange is conscious of both definitions and is making choices to deceive; the other is only aware of one definition and is being misled. The Government’s unsupported statement that there is no bad faith here seems premature.

³⁰ See, e.g., *Daubert*, 509 U.S. at 586-587 fn. 4 & 5 (detailing a portion of the persistent legal controversy following *Frye*, both prior to and in the wake of the adoption of the Federal Rules of Evidence).

³¹ See *id.* at 596-97 (disposing of objections raised by *amici*). Representatives of the scientific community, of whom *Daubert* broadly would come to divest authority to define what is “generally accepted”, argued the decision would stifle innovation. Curiously these arguments echo those published by the author to whom the term “junk science” is attributed, whose other notable contributions include alarmist works advocating for tort reform, see Peter W. Huber, *Liability: The Legal Revolution and its Consequences* (Basic Books 1990); a broadsided attack on claims to scientific truth, see Peter W. Huber, *Judging Science: Scientific Knowledge and the Federal Courts* (MIT Press 1999); theorizing that petroleum is an inexhaustible resource and that efforts to conserve energy are inherently counterproductive, see Peter W. Huber, *The Bottomless Well: The Twilight of Fuel, the Virtue of Waste, and Why We Will Never Run Out of Energy* (Perseus Books 2005); and advocating for the deregulation of the pharmaceutical and technology sectors to enable human experimentation, see Peter W. Huber, *The Cure in the Code: How 20th Century Law is Undermining 21st Century Medicine* (Basic Books 2013).

³² *Daubert*, 509 U.S. at 589.

scientific claims.³³

Performing this task, as the *Daubert* court suggests, is a meticulous judicial function. Entrusted to the court's sound discretion, questions of scientific evidence incorporate the whole of the FRE. Thus, the analysis necessarily encompasses relevance, probativeness and materiality, as well as the tendency of the evidence to mislead, prejudice and confuse.³⁴ Interpreting FRE 702 proceeds safely within principles of statutory interpretation³⁵, and begins with questions of what constitutes genuine science, knowledge, and scientific knowledge.³⁶ Rule 702 is clearly amenable to this approach³⁷, which is naturally concordant with the FRE as a whole³⁸, which in turn harmonizes with federal and constitutional law.

In place of this robust and integrated framework, the Opposition Brief offers up a

³³ While this is indisputably both the holding and the effect of *Daubert*, it has gathered force through subsequent repetition by the Court: “*Daubert*...held that the Federal Rules of Evidence *assign* to the trial judge the task of ensuring that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand.” *Kumho Tire Co., Ltd. v. Carmichael*, 526 U.S. 137, 141 (1999) (emphasis added, internal quotations omitted).

³⁴ See *Daubert*, 509 U.S. at 595 (discussing the intersection of Rule 702 and Rule 403); see, e.g., *United States v. Williams*, 583 F.2d 1194 (2d Cir. 1978) (discussing evidential reliability under *Frye*, but applying the FRE).

³⁵ *Id.* at 587.

³⁶ See *Daubert*, 509 U.S. at 588-90.

³⁷ The Court’s first cut at Rule 702, parsing the “scientific knowledge” clause from “will assist the trier of fact to understand or determine a fact in issue” is a helpful starting point, but bears further scrutiny. In this case, each clause can—and should—be broken down further. For example, because it is not transparent in how it produces results, FST cannot assist the trier of fact to “*understand facts*.” FST also produces a LR that is determined by a great deal of information that is not “at issue” in the case. The remarkable flexibility in the language of Rule 702 allows for such parsing as the facts call for here.

³⁸ The Court clarifies that “scientific reliability” is among the highest standards of legal reliability, forcefully demonstrating the point through statutory analysis. Noting its intention to employ the evidentiary, not the scientific, meaning of the term, the Court analogizes to Rule 602, which permits witnesses to testify to a matters only within their personal knowledge. The predicate requirement of personal observation “is a most pervasive manifestation of...*the most reliable sources of information*.” *Daubert* at 590, fn. 9 (*emphasis added, internal quotations and citations omitted*). Having indelibly linked evidentiary reliability to personal knowledge, the Court rejoins evidentiary and scientific reliability. *Id.* Far more than “not-junk science”, the Court’s authoritative interpretation establishes scientific reliability as *the most reliable source of information*.

new paradigm, one which isolates a phrase plucked from popular literature and injects it into the evidentiary analysis. It is an innovative proposal, suggesting that anything not condemned as “junk science”³⁹ is welcome⁴⁰, but it is simply wrong.⁴¹ While it is certainly true that a properly conducted *Daubert* inquiry would prevent “junk science” from being introduced, Rule 702 is hardly limited to preventing the admission of “junk science” alone. Substituting the inquiry’s floor for its ceiling is little more than rhetorical sleight of hand.

The Government would have every issue go to “weight”, and cross manage every scientific challenge. Cross examination may be a powerful tool, but the context matters. The scientific, technical, and mathematical arguments, coupled with those of software development, engineering, and validation must be evaluated at this pre-trial phase. “[C]ross-examination is inherently handicapped by the jury’s own lack of background knowledge.”⁴² This is particularly true when dealing with a complex, novel software like FST.

And of course, while we firmly believe FST is not entitled to such weight, there is abundant reason to be concerned that jurors will credit the DNA. Such evidence is particularly likely to prejudice jurors, who tend to afford such evidence undue weight. “DNA evidence is often viewed as being qualitatively different from other evidence because of its presumed scientific rigor and accuracy.”⁴³ Of course the Government urges the Court to admit this evidence: numerous studies have shown that jurors inherently trust DNA evidence and that they are more likely to convict in

³⁹ The single example of “junk science” ever defined by the Supreme Court, of course, is Phrenology. That is, expert “testimony...purport[ing] to prove a defendant’s future dangerousness based on the contours of [his or her] skull.” *General Elec. Co. v. Joiner*, 522 U.S. 136, 153, fn. 6 (1997). Attempting to take the Opposition Brief’s scheme seriously while also harmonizing *Joiner*, *Daubert* and Rule 702 is extremely difficult. Were *Joiner* to have limited *Daubert* as implied, beyond ensuring that Phrenology remains in history’s dustbin, Rule 702 would serve virtually no purpose.

⁴⁰ Gov’t Br. at 25, 29, 42, 53.

⁴¹ One will search *Daubert* and all the FRE in vain for the term “junk science.” It does not appear. See also *Amorgianos v. Nat’l R.R. Passenger*, 303 F.3d 256, 267 (2d Cir. 2002) (“To warrant admissibility, however, it is critical that an expert’s analysis be reliable at every step.”)

⁴² See *United States v. Glynn*, 578 F.Supp.2d 567, 574 (S.D.N.Y. 2008).

⁴³ See Joel D. Lieberman, et al., Gold versus Platinum: Do Jurors Recognize the Superiority and Limitations of DNA Evidence Compared to Other Types of Forensic Evidence?, 14 Psychol. Pub. Pol’y & L. 27, 32 (Feb. 2008).

cases where such evidence is introduced.⁴⁴

Much is also made of *Daubert*'s "liberalization" of the scientific evidentiary standards. Much repeated, this qualifier is often misconstrued, mistaking a method for an outcome. But *Daubert* never signaled the collapse of evidentiary integrity within the federal courts. It well may have strengthened them, affirming the authority of the district court as the gatekeeper. The source of the term "liberal" has three plainly-applicable sources that explain the persistent reference: First, in superseding *Frye*, which by its terms only applied to "novel" scientific claims, *Daubert* broadened the scope of what may be subject to admissibility challenge.⁴⁵ Second, *Daubert* affirmed the primacy of the FRE in deciding the admissibility of scientific evidence, and the FRE themselves liberalized and rationalized preexisting common law of evidence. Third, the term applies to the scope of the inquiry, which clearly is broadly permissive and not subject to a rigid or austere test. The court is empowered to act within its sound discretion.

Whether DNA testing in fact achieves what its proponent claims it achieves is a question that has branches in evidential authenticity and scientific validity of the technique from which the evidence was derived. Rule 702 permits the proponent to elicit expert testimony only when "based on sufficient facts and data"⁴⁶ and if "the testimony is the product of reliable principles and methods[.]"⁴⁷

The burden and standard of proof, and the presumption of innocence, already closely related and slippery concepts, are made more difficult to grasp when it is reinforced that the defense's "hypothesis" (a proxy for the accused's story, irrespective of whether or not he made an inculpatory statement) has failed an expert's mathematical testing. The Hd is imposed by the Government's lab.⁴⁸

While DNA's probative value might be expressed any number of ways, few means of

⁴⁴ See Kimberly Cogdell Boies, *Misuse of DNA Evidence is Not Always a "Harmless Error": DNA Evidence, Prosecutorial Misconduct, and Wrongful Conviction*, 17 Tex. Wesleyan L. Rev. 403, 406-7 (2011).

⁴⁵ Beyond the text of the FRE, *Daubert* plainly imposed no limit on the subject of what may be subject to an admissibility challenge. To arrive at the opposite conclusion, the Opposition Brief must ignore entire passages of the decision. See, e.g., *Daubert* 509 U.S. at 593, fn. 11 (explaining how *Frye* focused exclusively on "novel" techniques, but Rule 702 is not similarly limited).

⁴⁶ Rule 702 (b).

⁴⁷ Rule 702 (c).

⁴⁸ It is also dissimilar to affirmative defenses, which are received from common law, are waivable, and the defense exclusively owns the choice whether to employ them.

expressing evidential weight succeed to accomplish the goals set forth by NRC I.

Within the NRC framework, the signal means of expressing the weight of a single-source association has been the Random Match Probability (RMP). It answers the question: What is the likelihood of picking a profile at random from a given population and obtaining an association? It is relevant, in that it describes the rarity of the profile, which generally expressed in terms of odds (i.e., $1/10^{17}$), which often is a very remote probability indeed, typically satisfying the proponent. But while there is risk that the factfinder will misconstrue the meaning of vanishingly small numbers in the denominator⁴⁹, the RMP is a fair statement of rarity which, when presented accurately, scrupulously avoids burdening the defendant's presumption of innocence. It is a neutral expression of weight that is in every way consistent with NRC II's goal to conduct DNA testing "calculate probability of a match on the assumption that the suspect is innocent and the DNA was left by someone else."⁵⁰

C. FST poses an unacceptably high risk of false positives and exaggerated false associations.

No false positive testing was ever undertaken of FST. The reported false positive LR values were produced in the course of assessing the "distribution of LR expected when a suspect is actually not a contributor [to the evidence.]" This testing was performed with a set of mixed samples, comprised of combinations of profiles from only 14 contributors. Some of these 439 samples have since been shown to be repeats.⁵¹ Although the strikingly thin stock of source profiles in the comparison would skew the data, the programs authors only reported the aggregate number of mixed samples – 439 – to the DNA Subcommittee, continuing to conceal this information in a subsequent publication.⁵²

⁴⁹ Various errors of reasoning from statistics (named for the lawyers whose side would benefit from them) can always occur. With an RMP of $1/10$, the "chance of the defendant's guilt" remains, of course, unknown. The statistic only answers the chance of obtaining the profile when selecting from a given population. It neither asks *What is the chance that the defendant left the profile?* (Prosecutor's fallacy) nor *What is the chance that someone other than the defendant left the profile?* (Defense attorney's fallacy). William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence In Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 Law & Hum. Behav. 167 (1987).

⁵⁰ See NRC I at p. 29.

⁵¹ See Shapiro Aff., Def. Br., Ex.A. (Data concerning the racial characteristics of these 14 contributors was not retained.) Moreover, the 439 mixtures overstates how many *profiles* are compared, which is the key question. Not, how many times a technician hit *send*.

⁵² See Mitchell et al.

As reported to the DNA Subcommittee, the false positives appeared in the Executive Summary to the Validation Study. The lab attributed the aggregate false inclusion rate of 0.08% to allelic sharing (later changed by Adele Mitchell to 0.03%).⁵³ The aggregate false inclusion rate is not informative, particularly considering the thin composition of the mixed samples in the comparison. The same small group was being run against the same data set every time. In this context, the number of comparisons is not meaningful. The OCME could have continued to run more comparisons and driven their rate downward but it would not have made FST any safer or more reliable.

The lab did not present the important fact, which is present deeper in the Executive Summary: Nearly one in five of the comparison *samples* produced a false inclusion.⁵⁴ This figure more ably demonstrates the risk of false inclusion, not the number of times the same combinations are run.

The lab's statement that these false inclusions were produced by allele sharing begets further question. FST protocols are striking precisely for their notable lack of any failsafe to ensure against allele sharing. The lab's statements to the Subcommittee in this regard are contradictory, first suggesting that all false inclusions are because of allele sharing ("Due to allelic sharing, a small number of non-contributors, 0.08%, generated positive log LR's").⁵⁵ Later, it appears to be only the highest log LR that allelic sharing is the culprit: ("This non-contributor was fortuitously included in the mixture due to allelic sharing").⁵⁶ The implication of allele sharing in the sample comparison exercise was never further explored.

D. FST causes new constitutional harm in the guise of avoiding another.

First, the program explicitly removes the presumption of innocence from the defendant. To even formulate a relevant LR incorporating the accused, one must abandon the NRC I framework. It is otherwise impossible to "calculate [the] probability of a match on the assumption that the suspect is innocent and the DNA was left by someone else." Mathematical convenience is not sufficient grounds to take this step. FST can do it no other way.

Second, FST shifts the burden of proof onto the accused. He is required to say

⁵³ See Executive Summary, Def. Br., Ex. S, at 2.

⁵⁴ See *id.* at 12. ("Specifically, 82 (18.6%) of the 439 samples generated an LR ≥ 1 for at least one non-contributor.")

⁵⁵ *Id.* at p. 2.

⁵⁶ *Id.* at p. 12.

something to explain the evidence. As a matter of law, of course, this is simply false. The Government's similar contention – that if Mr. Johnson “believes there are other scientifically valid means of calculating an LR, he may perform those calculations and present his or her own LR as evidence at trial” – is another attempt to improperly shift the burden onto the defendant.⁵⁷ This is, of course, disallowed. The proponent of the evidence must answer for their evidence's shortcomings.

Third, Bayesian mathematics – variables, hypotheses, and outcomes – impose a false sense of dutiful rigidity on the criminal trial. For example, not only must the putative Hd explain the evidence, but it must do so in direct contest against the Hp. Certainly, from a given perspective, there is some utility in viewing the evidence and perhaps the entire case in such terms. The vice emerges, however, when this viewpoint is required. The defendant is never required to explain the evidence, even at his peril. He is within his rights to choose a theory that entirely defies the OCME's framework – that FST is error-prone, for example. The program is not capable of calculating a LR for that position, but it will still impose a Hd upon the accused just the same. Its false binary scheme distorts the truth of the law and the trial contest.

Fourth, FST lightens the Government's burden to prove its case beyond a reasonable doubt. By juxtaposing the Hp and Hd in a ratio, the statistic implicitly suggests equivalency. As with the way FST disrupts other aspects of the NRC framework, the highly precise language chosen for the standard of proof jury instructions bespeaks generations of wisdom passed down within the courts, imparted within the courtroom to attentive jurors. Scientists of whatever rank perhaps may not be faulted for lacking this structural insight. That does not justify its disruption.

Statistical formulations that do not fabricate or require the weighting of a putative defense theory, do not ask the factfinder to weigh the odds of the defendant's guilt, and which do not offer a score for the winner of the adversarial contest, have been a hallmark of an American legal orthodoxy precisely because they harmonize the rights, rules, law and interests at stake in criminal cases.⁵⁸ The burden and standard

⁵⁷ In support of this contention, the Government cites to a civil case, *In re: Gen. Motors LLC Ignition Switch Litig.*, 2015 WL 9480448 at *6 (S.D.N.Y. Dec. 29, 2015), in which the District Court held an expert's decision not to conduct in-person interviews was an acceptable methodology. Even in the civil context, there was no argument whatsoever that the other side had the burden to introduce alternative methodology; instead, the Court answered the same question before the Court here: did the proffered expert testimony meet the *Daubert* standard in its own right.

⁵⁸ See *State v. Skipper*, 228 Conn. 610 (1994) (holding admission of Bayesian statistic to prove element of crime was reversible error, as “it is antithetical to our criminal justice

of proof, and the presumption of innocence, already closely related and slippery concepts, are made more difficult to grasp when it is reinforced that the defense's "hypothesis" (a proxy for the accused's story, irrespective of whether or not he made an inculpatory statement) has failed an expert's mathematical testing. The Hd is imposed by the Government's lab.

III. The Government fails to demonstrate that FST is admissible.

A. The record of FST's scientifically unreliable defects and limitations.

FST is scientifically unreliable by design. While crystal-clear evidence demonstrates that it is not now the same formula or program as was once putatively subject to internal validation, nothing in its documentation supports the conclusion that such validation was ever effective for the purpose to which FST has since been put. Whistling past these defects, declaring they are irrelevant to admissibility⁵⁹, the Government embarks on an impressively dexterous defense of FST's inexcusable deficiencies. FST results are reliable, so it is claimed, even if the underlying science is not.⁶⁰

Abundant false framing of the Defense Motion fails to add a glimmer of clarity: Attacks on "the variables and assumptions of FST" still don't land on "the technique itself." FST, made in good faith and based in science, is thus reliable, regardless of the reliability of the statistic it reports.⁶¹

The sheer diversity rejoinders, inconsistent though so many may be, shows the trouble with FST. There isn't one good answer that is acceptable to the Government's position. There are, however, a few interesting ones: Serious flaws are ignored, dismissed, belittled, or – as with the matter of reliable statistics – confounded; others are skated over with surely feigned miscomprehension, passed off on the Subcommittee, misdirected, or relegated to the waiting bin of "weight, not admissibility." It is a remarkable performance, to achieve a perfect score for such

system to presume anything but innocence at the outset of a trial" and "courts must carefully guard against dilution of the principle that guilty is to be established by probative evidence and beyond a reasonable doubt."); *see also In re Winship*, 397 U.S. 358, 363 (1970).

⁵⁹ Gov't Br. at 1-2.

⁶⁰ *Id.*

⁶¹ Gov't Br. at 36. In apparent contradiction, the Government later argues that the *results*, need not be reliable after all, as the Federal Rules of Evidence don't reach the manner in which evidence is presented. Gov't Br. at 52. ("The manner in which OCME reports results, of course, is wholly irrelevant to FST's validity and admissibility."). We do not agree.

a demonstrably compromised artifact as FST.

Of course, to arrive there, the Government must materially misconstrue FST, a program that by design is particularly sensitive to the slightest change of “variables or assumptions”⁶²; twist the Defense Motion into a caricature; and alter the legal standard beyond recognition. In the enthusiasm, the Government failed even to address the Court’s January 6, 2017 Order granting a *Daubert* hearing.

At that hearing, the Government will have the burden of demonstrating FST’s evidential reliability. This will not be possible. We expect the evidence to show, generally consistent with the documentary record established thus far, that FST is scientifically unreliable, unfit for the purpose to which it has been employed, neither valid nor validated, and not otherwise admissible pursuant to the Constitution or the Federal Rules of Evidence. Some of the reasons are detailed in reply to the Government’s Opposition below.

1. Issues the Government concedes.

Whether implicitly, explicitly, or by complete omission, the Government concedes several issues, narrowing the scope of the contested issues for a hearing.

a. The formula of FST obtained under Protective Order is not the formula subject to the validation study.

Independent review of FST source code revealed behavior within the program had not previously been reported anywhere, and was omitted entirely from FST’s documentation.

The formula for FST obtained pursuant to protective order is not the same as that the OCME and the Government claim to have been validated. The Defense Motion asserted this issue.⁶³ To this the Government has been unresponsive.

Silence under these circumstances has been treated as an admission.⁶⁴ In fact, there can be no genuine dispute as to this question. The only troubling issue is why the United States fails to simply admit such an obvious fact. The programs are different. Whether it is admitted or the Court must so find – from presented evidence or from the Government’s tacit admission – we will seek such a finding.

⁶² *Id.* at 1.

⁶³ *See* Def. Br. at 45-46.

⁶⁴ *See Kule Rubin v. Bahari Group Ltd.*, 2012 WL 691324, at * 3 (S.D.N.Y. Mar. 5, 2012) (“Although a party may deny an allegation if the party lacks sufficient knowledge or information to form a belief about the truth of an allegation...in any other circumstance an answer that neither admits nor denies allegations is deemed to admit the truth of those allegations.”) (citations omitted).

The Adams Declaration demonstrates the basis for concluding the formula was altered. Different version numbers evidently assigned to the program appear on different FST reports.⁶⁵ Having inspected the source code, Adams saw no evidence of the use of a version control system, which “allows for modifications to be tracked.”⁶⁶ Discussing the implications of the lack of information, Adams lays out the applicable, non-binding guidelines for validation of software systems such as FST.⁶⁷ While the definitions with the standards lack precision, they demonstrate the value of re-validation in the face of changes.⁶⁸

He concludes the need to re-validate probabilistic genotyping software following changes of this nature is paramount: “Due to the complexity of probabilistic genotyping algorithms, seemingly small modifications to the code can have unintended consequences and “knock-on” or compounding effects that alter the intended behavior of the algorithms. The significance of a change should not be underestimated.”⁶⁹ Adams is “unaware of any background or validation materials pertaining to FST that describe the removal of loci from likelihood ratio calculations.” Nor is he aware “of any re-validation study conducted by OCME on FST since these changes to the FST program have been made.”⁷⁰ The lack of documentation of this change to the program – and when, where, how, and, perhaps, why it was implemented – must be addressed by the Government as the proponent of the evidence. This undocumented behavior – not to mention other behaviors that the program may conduct that are yet undiscovered – undermines the entire validation study, and the study’s review by the Subcommittee.

b. Locus-dropping behavior remains undocumented, unexplained.

The source code revelation about how FST truly works has not been explained. No documentation or other support for a claim that FST is in fact a validated program has been put forward. Clearing away the Government’s diversionary grumbling that

⁶⁵ Compare screenshots of FST: Appendix B to Adams Decl., Def. Br., Ex. C to Validation Study, Gov’t Br., Ex. U.

⁶⁶ See Ex C to Def. Br. at 5.4.4.

⁶⁷ Adams Decl. II at 3.6.

⁶⁸ *Id.*

⁶⁹ *Id.*

⁷⁰ *Id.* at 5.

which in itself amounts to nothing⁷¹, what stands un-contradicted? FST is performing a behavior in calculations that a) goes unreported when it is utilized, b) is not known to the OCME examiner herself, and c) skews results.

To these factual assertions, the Government is wholly unresponsive.

It has been further asserted that the alteration of FST's formula was d) not internally validated e) externally validated, and f) that no effort was made to seek approval for the alteration.

To these specifics as well, Government is unresponsive. No explanation, understanding or justification is therefore available.

2. Implications of FST being employed without validation or approval.

The locus-dropping function directly contradicts the Government's persistent and overt reliance on the validation study and DNA Subcommittee: neither serves admission here. Rather, both strongly favor exclusion, as the factors tend to demonstrate how unreviewable the workings of FST are, and how little oversight is in fact exercised over the OCME in fact. On this alone, the Court should have serious doubts about FST's satisfaction of the *Daubert* standard.

a. Eliminating data is only helpful to a hypothetical contributor.

Then, the claim that the behavior itself is simply discarding "no helpful information."⁷² This is false, but it is not further explained. There is no way for the OCME to make distinctions about what information would or would not be helpful in advance. The question begged is helpful to whom? OCME's answer – the *unlikely contributor*.⁷³ This breaks with long-established ethical and legal norms existing between forensic science and the law.

It is not only the identity of the beneficiary of the OCME's "helpfulness", but also the very fact of it. Loci drop from the LR not when the allele frequencies *equal* one, which is when they the locus would theoretically be of "zero" help. The loci drop at ■■■ which is still potentially important, quite helpful information. Especially considering the potential effect across multiple loci, as Nathan Adams observed with

⁷¹ See Gov't Br. at 50 (describing the undocumented, undetectable, highly influential behavior of a public laboratory's privately-held software employed to create evidence in criminal casework a "specter" of a problem).

⁷² *Id.*

⁷³ See discussion of Ethical Breach *supra*

the PenB sample.⁷⁴

b. Conflation with NRC II *theta* is misleading.

From misleading, the Opposition Brief then shifts to substantially conflate a wholly distinct principle: the *theta* correction from NRC II that happens to share the same numerical value of 0.03. It is noted that the response on this point strains credulity. The premise ignores the facts. FST the dropping of loci does assist to calculate a more conservative – here employing the standard, favoring-the-defendant definition as received from the NRC – estimation of population frequency. It is a separate function altogether.

In short, *theta* correction is about rounding estimates *up*. That way, the statistics might not be as probative, but they are biased in favor of the defendant – always. This is perfectly consistent with the NRC regime’s effort to ensure that all statistics comport with constitutional law.

Its purpose is to guard against under-sampling errors, and is generally addressed to the debated around population substructure. It always protects the defendant. It only corrects by adding to a basic floor minimum frequency, but it never subtracts data. Finally, when applied, *theta* applies to the entirety of the sample.⁷⁵

The undocumented behavior? Its purpose is mystifying. But it acts on *high* frequencies, not low. It appears to be concerned with eliminating data, not embracing it. It most certainly does not account for population substructure. It acts on one locus at a time, not on the entire sample. It cuts from loci with individual LR’s that are positive, and those that are negative.

The commonality appears to be that they are each other’s inverse.

Finally, it does not appear to behave the way it was initially described. For example, in a sample where one subpopulation (A) has a frequency estimate of [REDACTED] but the three others, (B), (C), and (D) each has a frequency estimate of 0.10, the behavior drops *all four* from the LR. Only (A) reaches the arbitrary cutoff, but the data from all four are lost.

To suggest that this is a spectral claim that concerned with “no helpful information”⁷⁶ is highly misleading.

c. Discredited hand calculations have no relationship to dropped alleles.

⁷⁴ See Adams Decl. II.

⁷⁵ NRC II at pp. 29-30.

⁷⁶ Gov’t Br. at 51.

Next, they revert to “manual” checks. This it has been urged on the Court before. Twice.⁷⁷ It has been thoroughly discussed. It has been discredited. Yet, the Government returns to the well: “Any irregularities in the program’s operation would have been exposed by this check.”⁷⁸ They were not. Mitchell’s claim is baseless. The hand calculations were far too cursory (not a complete model of the FST algorithm, but checked pieces of math), far too few (24 of 557,000), and in the main, not math.⁷⁹

The Government makes this unsupported statement overlooking the extensive discrediting of the so-called manual checks made by Mr. Johnson on May 27, 2016. The manual checks remain insufficient – in their overall number, in their lack of completeness, and included interpretations that were qualitative, not quantitative.⁸⁰ This argument is stale and wrong.

3. Documenting how hidden behavior eliminates data and the effect on the LR.

Nathan Adams has now demonstrated precisely how the undocumented behavior has an impact on case work.⁸¹ First, the operator is not informed that the loci are removed.⁸²

Adams used the “Pen B” sample from the validation study: In the validation study, the lowest LR reported for the Pen B comparison was 157. Using the version of FST provided to Mr. Adams, the lowest LR was 70.6.⁸³ This difference between the results expected pursuant to the validation study and that which was revealed using the version of FST provided to Mr. Johnson is of paramount importance. “FST is a deterministic system, which means that for any given set of input, it will produce the same output each time it is run.”⁸⁴ Mr. Adams points to the undisclosed dropping

⁷⁷ See Transcript of Oral Argument, May 9, 2016, at 6:20 (“but they checked all the math by hand with the very same formula that defense counsel has”), 42:21-43:11; *see also* Docket No.63 p.2.

⁷⁸ Gov’t Br. at 51.

⁷⁹ See Docket No.50, p. 13-15.

⁸⁰ *Id.*

⁸¹ See Adams Decl. II.

⁸² *Id.* at 4. (“I have not received any notice or warning to indicate FST has removed loci from its likelihood ratio calculations.”).

⁸³ *Id.* at 3.4.

⁸⁴ *Id.*

of loci as one explanation for the disparate results.⁸⁵ When he used 15 loci, the LR was 157 (the LR reported in the validation study); when he used 12 loci (by removing the three loci where, in isolation, the LRs resulted in 1), the LR was 70.6. Not only does this demonstrate that the undocumented behavior is occurring when one runs the version of FST provided pursuant to the Rule 17 subpoena, but it demonstrates how impactful it can be on the results. Because FST is in the business of creating ratios, it is impossible to discard individual elements without impacting the end result.

Mr. Adams explains that the skew of the results is unpredictable.⁸⁶ As an example, he cites that of the three loci that dropped by the program during his Pen B investigation, two (D13 and D16) have locus LRs greater than one and one (D3) has a LR less than one. The locus is dropped before the LR calculation is made, which leads to great uncertainty: “The loci removed from the calculation could entirely support the prosecution hypothesis; entirely support the defense hypothesis; or variously support the prosecution and the defense hypotheses.”⁸⁷ Moreover, Adams explained that because “[o]nly one occurrence of allele frequencies summing to \geq [REDACTED] must occur for a locus to be removed,” it is “difficult to predict” when the program will remove loci.⁸⁸ Mr. Adams observed “the removal of a locus from the likelihood ratio calculation for all subpopulations when any subpopulation’s allele frequencies sum to \geq [REDACTED] (emphasis in original). This means, the program might even remove a locus that supports the prosecution hypothesis for one subpopulation and the defense hypothesis for another.”⁸⁹ Such an inconsistency is of particular import when the OCME only reports the lowest LR across the four subpopulations.⁹⁰

He concludes: “Ultimately, the behavior of the version of FST I was provided deviates from the expected behavior of FST as described in its validation study.”⁹¹ This revelation directly undermines the Government’s reliance on the validation study and the Subcommittee’s review thereof to meet the *Daubert* standard. The source code review has shown how and why FST fails to meet the admissibility

⁸⁵ *Id.*

⁸⁶ *Id.* at 3.5.

⁸⁷ *Id.*

⁸⁸ *Id.*

⁸⁹ *Id.*

⁹⁰ *Id.*

⁹¹ *Id.*

threshold in federal court.

a. The drop-out rate was not empirically determined.

The OCME artificially lowered drop-out rates despite repeated claims that drop-out was “empirically determined.”⁹² This revelation came to light during sworn testimony given by Adelle Mitchell, the designer of FST at the OCME, that her testing was based on approximately 20 samples using data she never compiled nor turned over to the Subcommittee.⁹³ The Government’s curious response is merely to fault the Defense for utilizing the available record. This is wholly inadequate and fails to answer the merits.⁹⁴

4. The Government misunderstands the science.

a. The degradation module did not work.

Government scant treatment of the central issue of degradation disappoints. At the outset, it diverts, suggesting that FST works because the OCME “considered” degraded samples, at some point, to some degree, during the validation period. This is flat wrong. Consideration of degradation without incorporating it into the analysis is meaningless. To suggest otherwise would be akin to an absurd claim, like Volkswagen did not mislead about its diesel emissions because it was aware of the correct applicable environmental standards.

Here, however, the chapter of documentation the OCME amassed discussing the degradation module firmly establishes the lab’s knowledge that FST was incapable of managing samples subject to even moderate levels of exposure to real-world conditions. This directly undermines FST’s validity for employment with real-world cases like Mr. Johnson’s, the very purpose to which it has since routinely been put.

The government then misrepresents the evidence, claiming that OCME abandoned the degradation module simply because “degraded samples produced no better results than those derived from higher quality samples.”⁹⁵ That is not what the OCME says. According to Mitchell, the degradation module failed to separate the LR’s of true contributors from those known to be false:

Ultimately, it was determined that, in general, use of the degradation module as programmed resulted in LR’s closer to 1.0 for both true

⁹² See Def. Br. at 17.

⁹³ *Id.*

⁹⁴ Gov’t Br. at 39.

⁹⁵ Gov’t Br. at 39.

contributors and non-contributors. That is, this approach did not increase the overall separation between true contributors and non-contributors (data not shown). This is an area in which further study is warranted, as improvement in quantification of degradation or identification of moderately to severely degraded samples coupled with changes to the degradation model might improve performance.

The degradation module reveals FST's flaws, how much it relies upon pitch-perfect data. But, as with so many other stages of the putative validation, the OCME turned its blame on the data, not the tool.

Finally, the government obscures, ironically using language suggesting the opposite by alleging that the lab presented the degradation module results "transparently" to the October 8, 2010 meeting of DNA Subcommittee. The publically-available minutes of this meeting fill four lines. Also on the agenda of that meeting were several other issues. There is no doubt that OCME sought to have FST endorsed for use. This was not a neutral evaluation with full access to the data, but rather a PowerPoint presentation performed by an interested party.

The question for this Court, is whether this passes for scientific study. It most certainly would not suffice for pharmaceutical or biomedical research. It should not here, where grave issues of liberty are at stake.

b. Threshold obscures valuable data.

As described in Mitchell, et al., the Drop Study was conceptualized to yield the maximum amount of information. This choice seems sensible, and depending on the object of the study, is not inherently problematic. But every such choice altered testing conditions beyond where they were reasonably related to the kind of forensic casework to which FST would be employed: With low-quality, likely degraded, mixed samples of unknown origin. The validation question is whether FST is suited for that purpose. In the Drop Study, enhanced inputs helped keep the program running smoothly, as the vast majority of samples were drawn from pristine blood or buccal swabs. Unusually sensitive materials and equipment gave consistently clear readings. Moreover, the lab employed special protocols that were reserved solely for the Drop Study but are not used in cases, such as where, in order to correct for the wide uncertainty in the quant process, each mock mixed sample was quantified nine times – many multiple times more than case evidence ever is.

Perhaps most emblematic of the Drop Study's artifice, the lab considered alleles below the 75rfu detection threshold in the Study, while such information is forbidden from casework. In Mr. Johnson's case, simply considering alleles below 75rfu would dramatically impact the LR, potentially bringing in so much genetic information as to invalidate FST's use. At minimum it would greatly reduce the misleading LR

FST created.

As he learned of the growing list of problems with the Study, Dr. Ranajit Chakraborty – who himself had seconded the motion to endorse FST for use in casework – became concerned about the program’s suitability for its purpose.⁹⁶ His list of objections to the OCME’s methods is extensive, detailed, and well-informed. But his conclusion that their impact is that FST is not suitable for “real world” forensic work is notable, as it echoes the emerging guidelines for validation of probabilistic genotyping systems, a category the OCME would claim for FST.

How the Drop Study was conceived and implemented keep it from having a true relationship to forensic casework. The Study validates FST as much as wearing a raincoat indoors: It is related to the purpose, but reliable conclusions can be drawn from the available data. Whether it was at the design stage; when data results began to accumulate and showed virtually no instances of drop-out had occurred; or at some other moment during the Drop Study—the OCME eventually became aware that the Study wasn’t lining up with casework.⁹⁷

5. The Subcommittee was an inadequate stop gap.

a. Simulated alleles were obscured from the Subcommittee.

The Government identifies two pages of the validation study that reference the fact that “genotypes were simulated at these loci for the Black, Caucasian, and Hispanic population samples” and include two tables of said frequencies.⁹⁸ Yet, it is found nowhere in the Mitchell article⁹⁹ It also was not brought to the attention of the Subcommittee. The Government does not contend otherwise. Moreover, the

⁹⁶ See Declaration of Dr. Ranajit Chakraborty, Dkt. No. 24, 15 Cr. 153, (S.D.N.Y.), Attached at Exhibit B; Declaration of Dr. Ranajit Chakraborty, Dkt. No. 29, 14 Cr. 414 (E.D.N.Y.), Attached at Def. Br., Ex. O.

⁹⁷ See Def. Br., Ex. M (Adele A. Mitchell, Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop in, 6 Forensic Sci. Int’l: Genetics 749-761, 756 (2012)) (“Mitchell et al.”) (“The samples used were pristine buccal swabs. Results may differ if samples displaying more of the complicated characteristics of many evidence samples. Further study of these phenomena is warranted.”)

⁹⁸ Gov’t Br., Ex. U.

⁹⁹ Gov’t Br. at 42; The Government overstates the case for Mitchell et al., as a “peer-reviewed validation paper.” The article was not a “validation paper” in the sense the Government seems to intend, in that it had nothing to do with the process of FST’s putative validation. It is a *record*, or a *recounting* of the Study, published approximately two years after FST was approved for coursework.

validation study on its face creates more obfuscation than clarity. “Because the frequencies must add to 1.0 in the simulation, all allele frequencies were reduced slightly to accommodate this.”¹⁰⁰ That statement is wholly without support: where did the simulated numbers come from? Is that source reliable? What is “slightly”? What scientific principle allows an amorphous adjustment? This is not a valid application of scientific principles.

IV. The Government’s *Daubert* factor analysis fails to support FST’s admissibility.

The Government addresses five factors under the *Daubert* rubric. As discussed above, the list is neither mandatory nor exclusive.

A. Factor One: Testing.

The Government dispenses with the first factor by citing the length of the OCME’s initial internal validation study and the oversight of the Subcommittee. The validation study conducted by OCME of the software it created internally was conducted internally. No independent testing was ever undertaken. Indeed, to safeguard their proprietary interest in the software, the OCME fought hard to prevent its release even after a Rule 17 subpoena was issued for it, with consent for a protective order.¹⁰¹ However, OCME then put FST to use without even a minimal case study, case-by-case performance testing, or sufficient false-positive testing. The deficiencies laid out raised in Mr. Johnson’s brief demonstrate the OCME’s failure to sufficiently validate FST.

In particular, there are serious questions raised whether the version of FST that was used in Mr. Johnson’s case – 2.5 – was ever validated.¹⁰²

B. Factor Two: Peer review.

The Government cites to the DNA Subcommittee of the NY State Commission on Forensic Science’s approval of FST; ASCLD-LAB, a non-profit professional society of crime laboratory directors and forensic science managers, which accredited OCME generally, not FST; and the National Forensic Science Technology Center (NFSTC)¹⁰³, a . These organizations are not academic institutions. The *Daubert*

¹⁰⁰ Gov’t Br., Ex. U, p. 163.

¹⁰¹ See Docket No. 63, Letter from OCME, at p. 2-3.

¹⁰² See Government’s concessions, *supra*.

¹⁰³ That the Government would cite to the NFSTC exemplifies its misconstruction of *Daubert*’s holding, as it conflates “peer review”—which has an established meaning and purpose that is related to establishing and maintaining the highest standards of scientific

Court contemplated peers as a “competitive, unbiased community of practitioners and academics.”¹⁰⁴

The DNA Subcommittee in a unique body, formed by statutory mandate.¹⁰⁵ The Subcommittee does not operate wholly independently of the laboratories and processes that it reviews. Dr. Charles Hirsch, while he was the Chief Medical Examiner at the NYC Office of the Chief Medical Examiner, was a member of the Subcommittee himself.¹⁰⁶ Drs. Caragine and Mitchell were in attendance, as were many other employees of the OCME, during the meetings at which FST was discussed.¹⁰⁷ The questioning of Drs. Caragine and Mitchell by the Subcommittee members reveals significant deference to the architects of FST.

Additionally, the Subcommittee’s “review” of FST took place as a portion of 5 meetings of the Subcommittee. Each meeting, covering all topics, lasted approximately 3 hours. At some meetings, as few as 5 members of the Subcommittee were in attendance. The Subcommittee does not appear to have engaged in any independent investigation; their questions were posed to Drs. Caragine and Mitchell, but there appears to be no follow up. (One Subcommittee member abstained on the ground that he didn’t fully understand what FST entailed.)¹⁰⁸ An accurate evaluation of the Subcommittee’s review of FST falls short of the

rigor—and a laboratory audit, which is at best only tangentially related to FST at all. The general audit of the OCME’s lab would likely encounter FST as part of ensuring that lab procedures comply with ASCLD-LAB standards for documentation, but would be extremely unlikely to be designed around *testing FST*. The company NFSTC, does not hold itself out as having any expertise in probabilistic work. It does, however have a very strong business in government contracts as a “trusted provider” working with laboratory leadership. It is a multi-million dollar non-profit company specializing conducts laboratory audits for hire. As compared with the rigors of peer review, in which one’s work is submitted and judged on the merits, an NFSTC audit is a collaborative process. The audit team meets with laboratory leadership initially, and before finalized, shared a draft report is shared that “will not be finalized until completion of the review process.”

See NFSTC Preparing for Your Audit, Attached at Exhibit D. This is neither rigorous nor independent, but the collaborative service of a “trusted provider.” [//www.nfstc.org/wp-content/files//2015_FORM_990_-_NFSTC_Final.pdf](http://www.nfstc.org/wp-content/files//2015_FORM_990_-_NFSTC_Final.pdf) at p.33.

¹⁰⁴ See *United States v. Starzecpyzel*, 880 F. Supp. 1027, 1038 (S.D.N.Y. 1995). (This reasoning also disqualifies the Subcommittee from serving as “peer review.”)

¹⁰⁵ N.Y. Exec. Law § 995-b

¹⁰⁶ Gov’t Br. at 19.

¹⁰⁷ Gov’t Br., Ex. J.

¹⁰⁸ See Subcommittee meeting, 12/7/10.

descriptor “exhaustive” put forth by the *Rodriguez* Court.¹⁰⁹

Moreover, these certifications are somewhat circular: ASCLD requires, among other things, validation of techniques employed by the lab; the DNA subcommittee, which is credited with validating FST, ensures that labs meet the accreditation standards set forth by ASCLD. The FBI Quality Assurance standards also require a lab to be accredited by a non-profit, like ASCLD.

The DNA Subcommittee members engaged in a debate about the worthiness of ASCLD’s “accreditations” and “audits.” A member expressed significant concerns about lab scandals involving labs that had ASCLD accreditation (calling it a “very bad record”) and cited the over-reliance on ASCLD (reasoning they were “the only game in town”).¹¹⁰ The skeptics on the Subcommittee draw a contrast between the “audits” that ASCLD performs – where the subject of the audit chooses the cases the auditor reviews and is given notice of the audit – and audits in other industries (like hospitals) where the external body appears unexpectedly and chooses the items to check.¹¹¹ The Subcommittee agreed to have the head of ASCLD appear before the Subcommittee and answer pre-submitted questions addressing these concerns.¹¹²

C. Factor 3: Known or potential error rate.

The Government argues, confusingly, that the error rate factor weighs in favor of reliability because there can be no known error rate. The brief fails to even cursorily explain how the latter fact (unknown rate) supports the former conclusion (favor of reliability).

Then, the Government proceeds to conflate the lack of a published error rate with the false positive study. As discussed above, the rate of false positives is also

¹⁰⁹ See Gov’t Br. at 36 (citing *People v. Rodriguez*, Ind. No. 5471/2009).

¹¹⁰ See Subcommittee meeting, 12/7/10 at 1:58:00.

¹¹¹ *Id.*

¹¹² Some would argue that the standards are so broad that they are no standard at all. In operation, rather than to ensure quality or truly police the industry, that the standards function solely to provide enough of an appearance of regulation to allow for the evidence to meet admissibility requirements. The steady persistence of quality control problems at forensic laboratories across the country, including the OCME, adds weight to this critique. Absent a federal commitment to reform forensic practice through the imposition of binding regulations or passage of law, it is difficult to envision how a national organization such as SWGDAM could police the industry more effectively.

unknown.¹¹³ The Government contends that .03% of the more than 500,000 comparisons were false positives.¹¹⁴ This demonstrates a complete misunderstanding of the false positive study. Mr. Johnson's brief details the limitations of the study: 14 contributors, skewed and unknown racial characteristics, etc.¹¹⁵

D. Factor 4: Protocols and review.

The Government relies on the existence of the OCME's FST technical manuals to support its contention that FST testing is subject to stringent protocols and review. The protocols, on their face, are notable for their overbreadth ("the analyst's discretion should be used when doing this determination")¹¹⁶ and vagueness ("Samples with too few or too many alleles should not be interpreted or used for comparison").¹¹⁷

Yet, when Mr. Johnson identified a significant failure to follow the protocols in his case – the incorrect injection times¹¹⁸ – the Government's response was to claim that the protocols are simply advisory: they "simply say that the analyst 'may' use particular parameters, not that the analyst must use those parameters, belying the defendant's claim that the analyst violated 'clear directives.'" ¹¹⁹ In other words, the Government believes the protocols are not stringent at all. If the Government's position is that the protocols need not be followed, then the existence of the

¹¹³ See Shapiro Aff., Def. Br., Ex. A., at ¶ 3 ("Due to design flaws in the "false-positive" performance checks, the rate of false-positives produced by FST is unknown.")

¹¹⁴ Gov't Br. at 34.

¹¹⁵ See Def. Br. at 28-30. See also Mimi Mairs, Special Counsel, Forensic Biology at OCME stating in response to court's inquiry about availability of racial data underlying the FST study: "Doctor Mitchell states that she and Doctor Carajine [sic] tallied up on a piece of scratch paper and that piece of scratch paper was not saved so when I say that there is no document much less a formal document, there is none." Mairs testimony, Collins Transcript, 6/7/2013.

¹¹⁶ OCME Protocols at p. 348.

¹¹⁷ *Id.* at 345.

¹¹⁸ See Def. Br. at 47-49.

¹¹⁹ Gov't Br. at 52. The Government does not challenge the fact that the injection parameters "were injected" as the table indicates in the validation study used high (IR) for samples 200 pg and below. See Def. Br., Ex. S. at p. 6. If the OCME protocols differ from that which was tested during the validation, the Government has uncovered a new reliability problem. (If the protocols do, in fact, mandate compliance with the injection levels from the validation study, Mr. Johnson's samples did not conform to them.) Def. Br. at 48-49.

protocols does not increase the reliability of FST.

The insufficiency of the protocols was demonstrated by the circumstances and investigation that led the then-Deputy Director of the OCME, Theresa Caragine, who participated in the creation of FST, to resign. Her resignation followed an investigation into the OCME's Department of Forensic Biology by the New York State Inspector General's Office (OIG).¹²⁰ OIG was designated by the New York State Commission on Forensic Science to serve as an independent investigatory entity. Dr. Caragine was found to have "ignored laboratory protocol regarding resolution of disputes."¹²¹

Specifically, Dr. Caragine was disciplined for disagreeing with another criminalist about the number of contributors to a mixture in a case where DNA was collected from the swab of a gun. She changed the analysis from another criminalist's conclusion that it was a two-person to her opinion, which was three-person mixture; she changed the report with the results accordingly. OIG found that she did not follow the appropriate dispute protocol (while taking no position regarding the number of contributors in the mixture). As the dispute continued, "[o]f particular note, the suspect report in the case at issue was rewritten six times; yet, after each revision, the computer program utilized by the OCME overwrote the previous draft and only the most current draft remained. As such, the reader of the final report would be ignorant of the dissension among the criminalists and Caragine regarding this case."¹²² After being confronted with another two instances of reassigning case work when she disagreed with the criminalist's findings, she resigned.

The OCME's overbroad protocols leave casework to discretionary calls. Teresa Caragine's cases are not outrageous so much because she exercised authority, but because she was not prevented from doing so by clear Standard Operating Procedures. The most alarming issues in the OIG report may be as the way different analysts reached opposing conclusions for which FST was able to produced quantitative support. Whether the protocols are treated as rules, guidelines, or the operating instructions for lab equipment seems to be in the eye of the beholder. The factor must weigh firmly against scientific reliability and admissibility.¹²³

¹²⁰ See Office of the Inspector General Report, December 2013 ("OIG Report"), Attached at Exhibit C.

¹²¹ *Id.* at 24-25.

¹²² *Id.* at p. 34.

¹²³ The OIG also found that "a substandard criminalist worked that the OCME for over 10 years despite consistent poor reviews and while significant errors when undetected." *Id.* at p. 14.

E. Factor 5. Scientific community.

The Government again misconstrues this factor, impermissibly substituting the legal community for the scientific community. Stated bluntly, the general acceptance inquiry is directed towards scientists, not courts. The standard concerns the debate or controversy – conversely the consensus or general acceptance – within the relevant scientific community. As the *Frye* court phrased it, “in the particular field in which [the evidence] belongs.”¹²⁴ Neither the *Frye* court nor any subsequent understood the relevant scientific community analysis to include a head-count of judges.

Of course, in the process of deciding any matter, a court in the sound exercise of its discretion may consider what weight to assign the non-binding decisions of sister courts applying a different legal standard in a markedly different system. The overwhelming number of cases compiled by the government – most unpublished – cannot provide a fresh perspective on FST, principally because only two courts conducted evidentiary hearings: *Collins* and *Rodriguez*. Of these, only the *Collins* court published its opinion.¹²⁵

Though the government suggests the state record is dispositive, a federal court must rest upon an independent evaluation of the evidence. Caution is warranted, moreover, when considering the New York State record. The state’s employment of a *Frye* standard is one which is far less stringent than *Daubert* or the Federal Rules of Evidence call for.

Further, there is peculiar difficulty with a public crime lab using a proprietary tool in *Frye* jurisdiction. The test calls for general acceptance in the “relevant scientific community.” This factor is never self-defining. Exaggerating to illustrate the point, the government would tend to define the relevant community as forensic examiners testifying for the prosecution; while the defense would prefer a group of confirmed skeptics with sterling credentials. In reality, however, the OCME has captured the market of FST users. Refusing to release the program even under protective order for independent assessment, the lab ensures that the relevant scientific community

¹²⁴ *Frye v. United States*, 293 F. 1013, 1014 (1923).

¹²⁵ The *Rodriguez* court, as the Government cites at Gov’t Br. 35, rested heavily on the acceptance in the scientific community of PCR-STR DNA testing and likelihood ratios accommodating allelic drop-out and drop-in. Gov’t Br. at 35 (citing *Rodriguez* at 28). In doing so, the *Rodriguez* court did not address the acceptance of FST’s application of those principles.

is limited to employees of the OCME – hardly an objective cohort.¹²⁶ Drawing experts from so narrow a pool as has the OCME, the conclusion that FST is generally accepted is predictable. As a practical matter it is impossible for a New York State trial judge to exhaustively conduct a *Frye* analysis.

V. Conclusion

FST was retired from new case work by the OCME on January 1, 2017. The proprietary software the OCME had created and used exclusively – after a multiple-years long formation process and complete internal overhaul – was forsaken in lieu of a newer software on the market. The new software came at a significant cost to the OCME for both the program itself and preparing the laboratory and training the staff to adapt to new software. In other words, OCME’s decision could not have been made lightly.

Yet, the Government frames the transition simply: they say the OCME is replacing its technology “because of changes” to the FBI’s Combined DNA Index System (CODIS) database.¹²⁷

That change was, in fact, a decision by the FBI to demand more thorough DNA testing: to run comparisons at 20 loci instead of only 13 loci. The FBI created a more rigorous threshold for testing – and demanded that labs such as the OCME follow suit. FST would not or could not satisfy the more rigorous standard.

At the heart of *Daubert* lies a central question of reliability. Here, the scientific community, led by the FBI, has reached a new level of reliability in testing, and OCME has officially decided to leave FST behind. For this reason – and the many others laid out in this brief and in Mr. Johnson’s initial challenge – we ask that you conduct the hearing ordered by this Court and ultimately, grant this motion *in limine*.

Respectfully submitted,

/s/

Christopher Flood, Esq.

Sylvie Levine, Esq.

Robert M. Baum, Esq.

Counsel for Kevin Johnson

¹²⁶ *But cf. Case of Canavan*, 733 N.E.2d 1042, 1053, n.6 (2000) (requiring a relevant scientific community to “be defined broadly enough to include a sufficiently broad sample of scientists so that the possibility of disagreement exists”).

¹²⁷ Gov’t Br. at 21.